Multi-timescale analysis of sequential behavior decisions in fly grooming.

primoz.ravbar@gmail.com lab website: https://labs.mcdb.ucsb.edu/simpson/julie/research

*PRIMOZ RAVBAR[1], KRISTIN BRANSON[2], JULIE H SIMPSON[1]

[1]UCSB [2]Janelia HHMI

Introduction: Fly grooming behavior is composed of variable sequences of discrete and continuous actions and can be used as a model behavior for studying sequential decision-making. To model the grooming sequence with suitable statistical power, we needed to collect and annotate very large amounts of video data. We overcome this challenge with a new method for automatic behavior recognition. The method is robust in recognizing types of grooming behaviors in freely moving flies across various experimental conditions and where parts of fly's anatomy are occluded. The system is based on extraction of rotation/translation invariant spatio-temporal features that can be used for classification supported by unsupervised learning algorithms.

Figure 1 Tracking of the fly in the arena

First we locate the fly in the arena. This is done by movement alone (by accumulating the differences between neighboring frames). We can reliably track the animal during ~30 minutes of video.

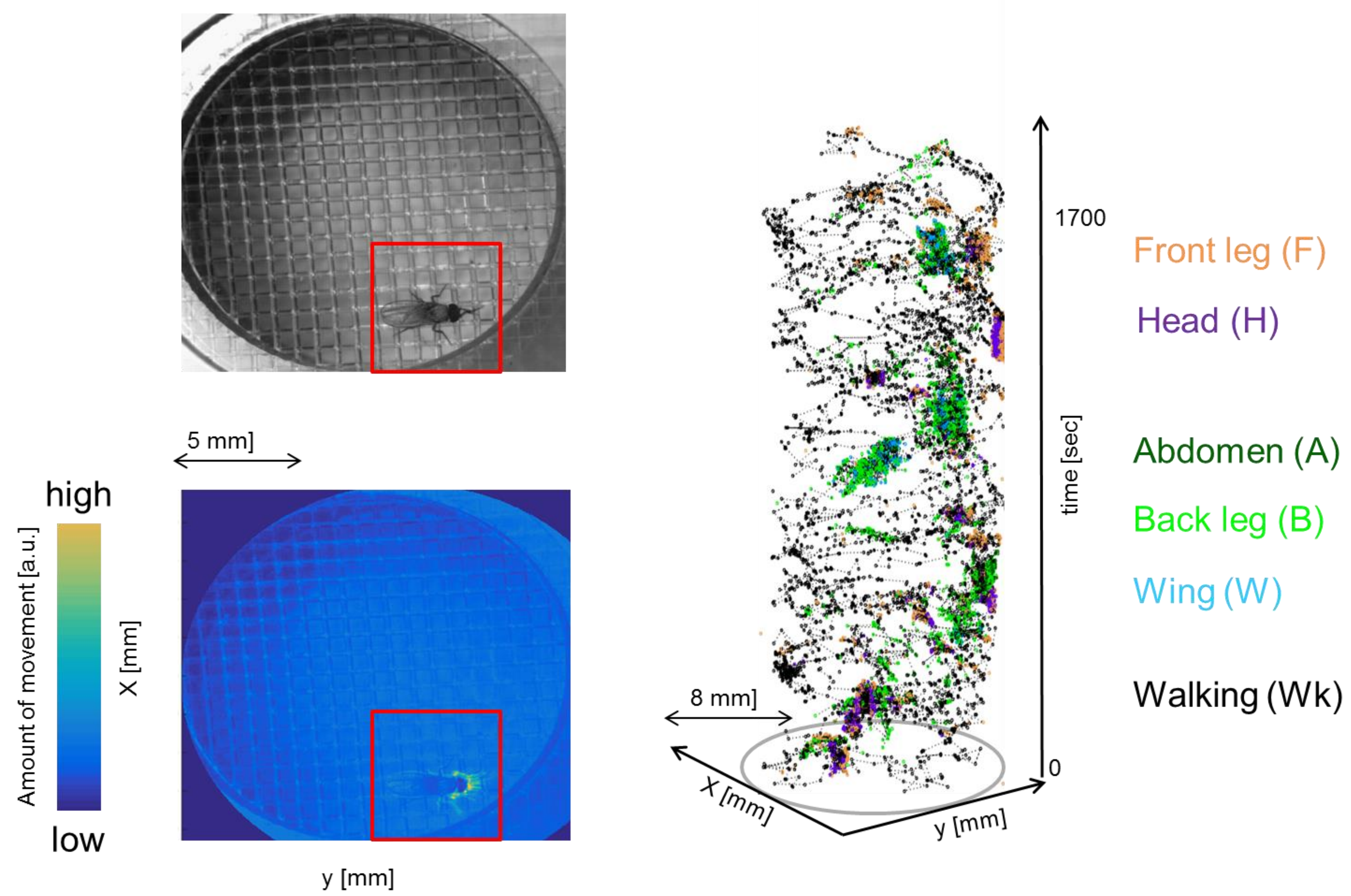


Figure 2 Spatio-temporal images

Next we extract spatio-temporal features from a stack of frames (.5 sec). Left: Time-traces of light intensity in each pixel (blue and red time-traces of blue and red pixels) are used to compute spectra by FFT. Mean frequency of the spectra is computed. Pixels in which light intensity changes periodically tend to have higher mean frequency of their spectra (red time-traces, blue traces are less periodic). Right: Spatial-temporal image (ST-image) is produced from the stack of frames, where each pixel is assigned the value of mean frequency of its corresponding spectrum.

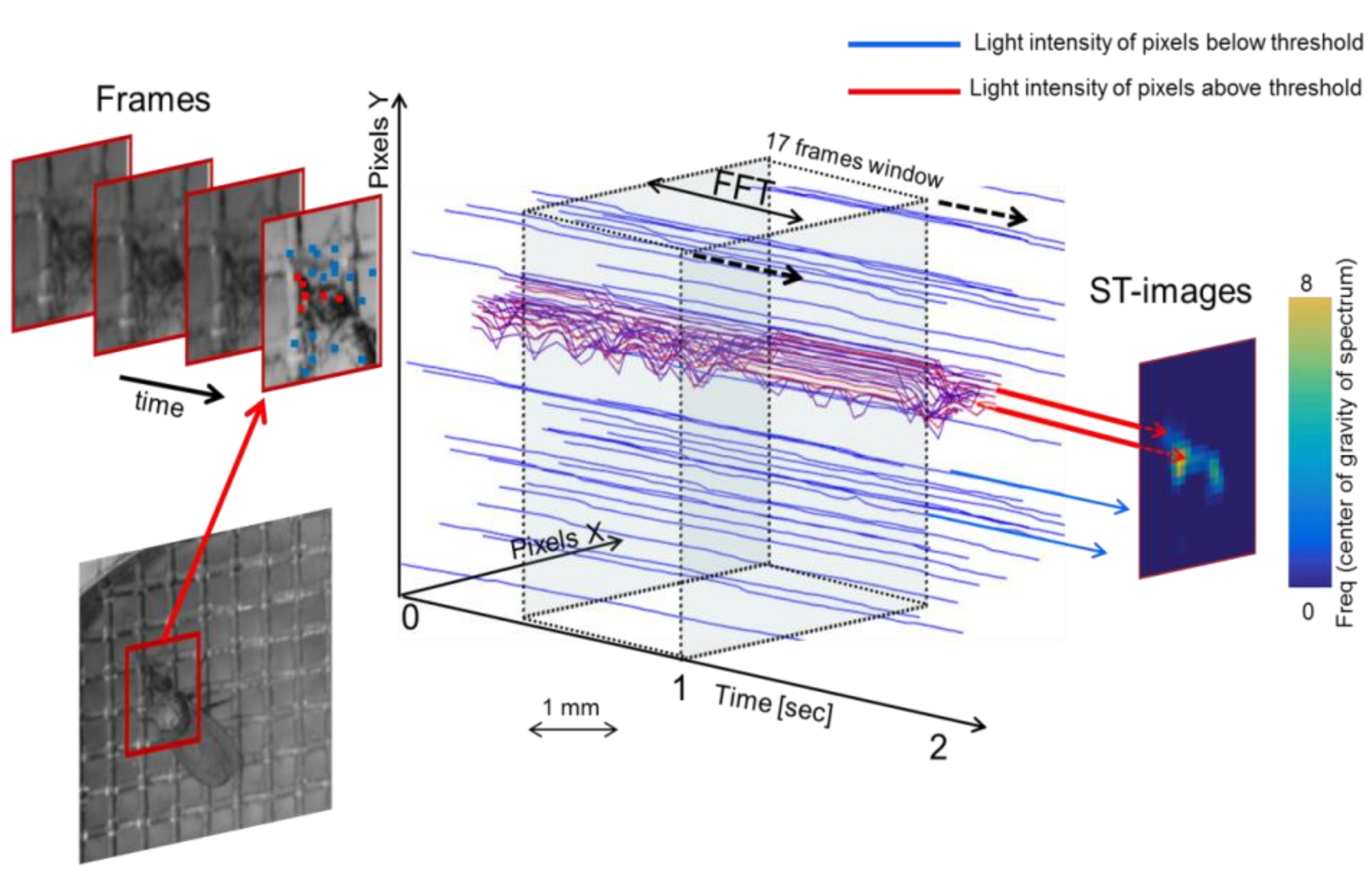
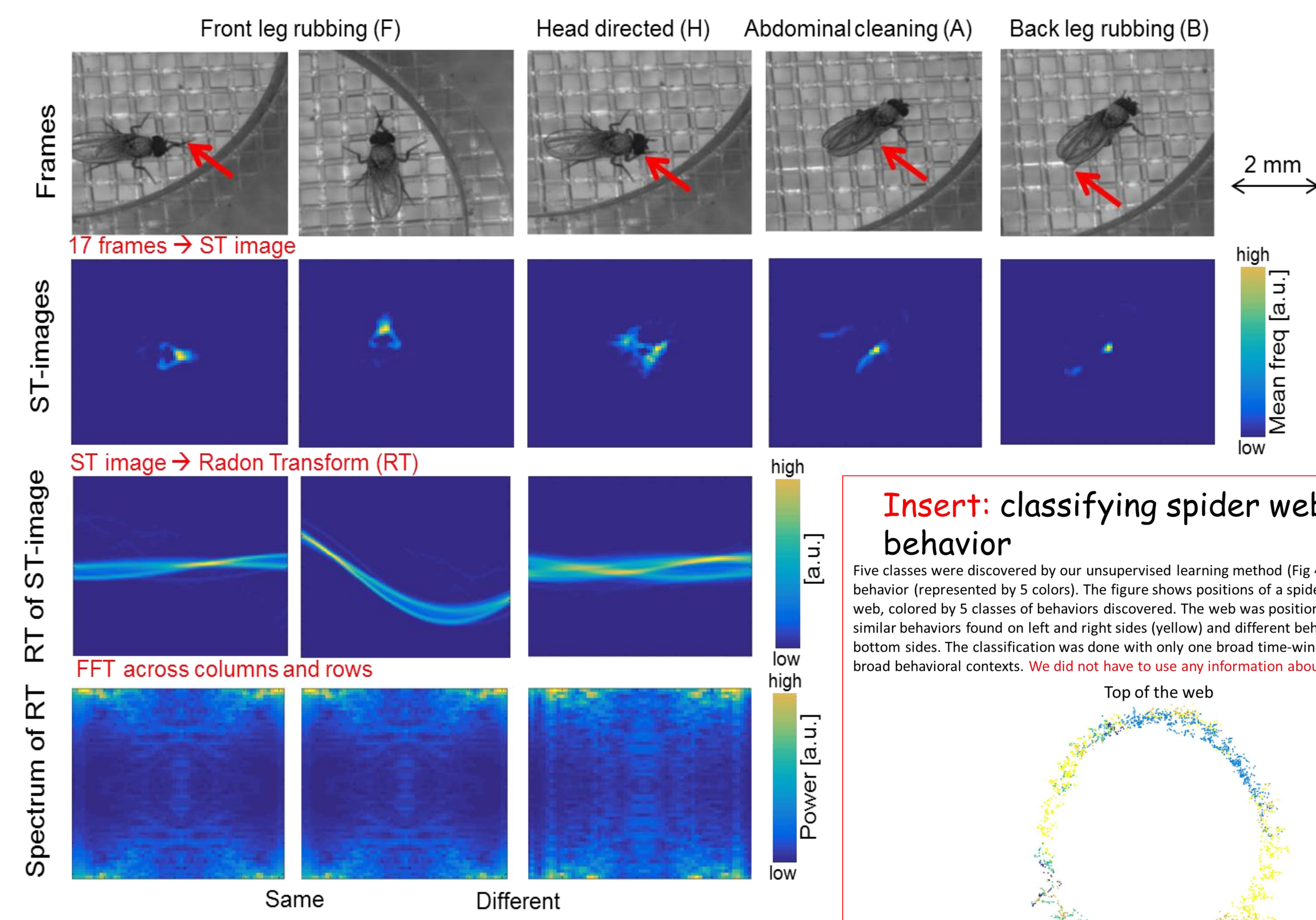


Figure 3 Rotation-translation invariance

ST-images represent "shapes of behaviors". We use Radon Transform to make these shapes rotation and translation invariant. The resulting images (spectra of RT) contain information about the shape but are rotation-translation invariant. They are used as input for the unsupervised learning. Examples of four behaviors are shown (top row). First and second behaviors are the same (first two columns). The corresponding ST-images of the behaviors (second row) are transformed into rotation-translation invariant images (bottom row). These images are identical for the same behavior (first two images).



Insert: classifying spider web-building behavior

Five classes were discovered by our unsupervised learning method (Fig 4) in spider web-building behavior (represented by 5 colors). The figure shows positions of a spider on the outer rim of the web, colored by 5 classes of behaviors discovered. The web was positioned vertically. Note the similar behaviors found on left and right sides (yellow) and different behaviors found on top and bottom sides. The classification was done with only one broad time-window step (see Fig 5) to find broad behavioral contexts. We did not have to use any information about spider's anatomy.

Video data collected by Andrew Gordus

Figure 4 Unsupervised learning

Randomly sampled videos are used as a training data-set for unsupervised learning (dimensionality reduction) by applying Singular Value Decomposition. Left: Each rotation-translation invariant image (containing 6400 dimensions – pixels) can be decomposed into 30 dimensions (bases/filters). Other dimensions are discarded. From the training data-set we retain 30 bases. For each new data point (new image) we compute projections to the 30 bases, thus compressing each 6400-pixel image into 30 values – final spatio-temporal features. Middle, bellow: example of 30 spatio-temporal features from 50 sec of a movie, aligned with a human-labeled ethogram. Right: t-SNE decomposition can be used to visualize how behaviors fall in different regions of the original 30 dimensional space, suggesting that these 30 features are sufficient to separate the behaviors from each other.

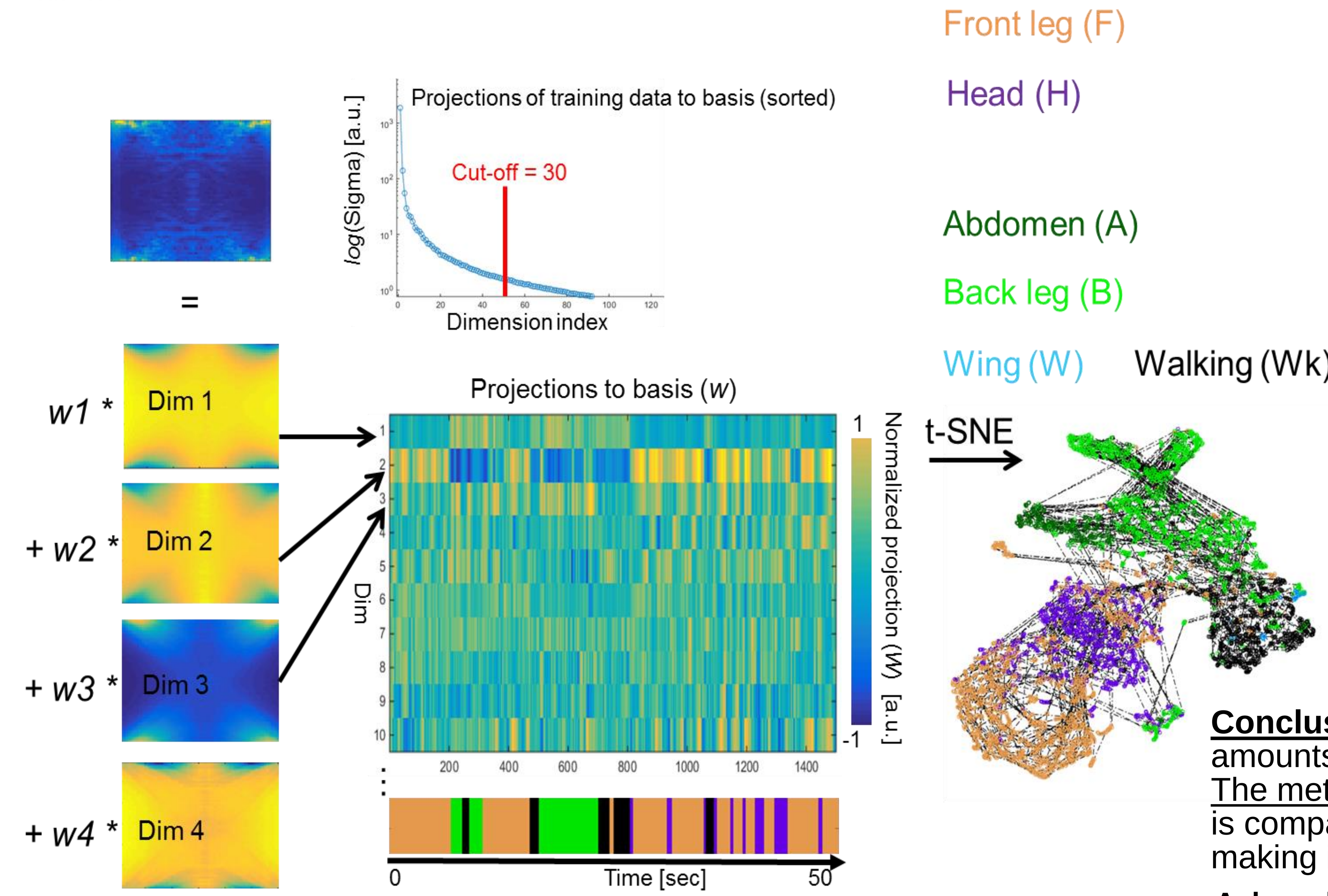


Figure 5 Two-timescale classification and validation

Spatio-temporal features are classified by Gaussian Mixture Modeling in two separate steps: first broad time-window ("widowed features") is used to identify behavioral context (Anterior or posterior grooming, or walking). Then each behavioral context is classified separately into individual grooming behaviors. The two-step version reduces the major source of ambiguity, between front and back leg rubbing (F, B). Left: Human-labeled (H) and machine-labeled (M) ethograms are shown below the spatio-temporal features, for each step (note the correspondence between features and behavior). Right: for each step, examples of the same data as it falls in 3-D space are shown. Colors correspond to behaviors identified by humans and the machine (H,M). Confusion matrices are shown below. The agreement between the machine and humans is comparable to the agreement between different human observers. Left, below: ethograms from 93 dust-stimulated flies.

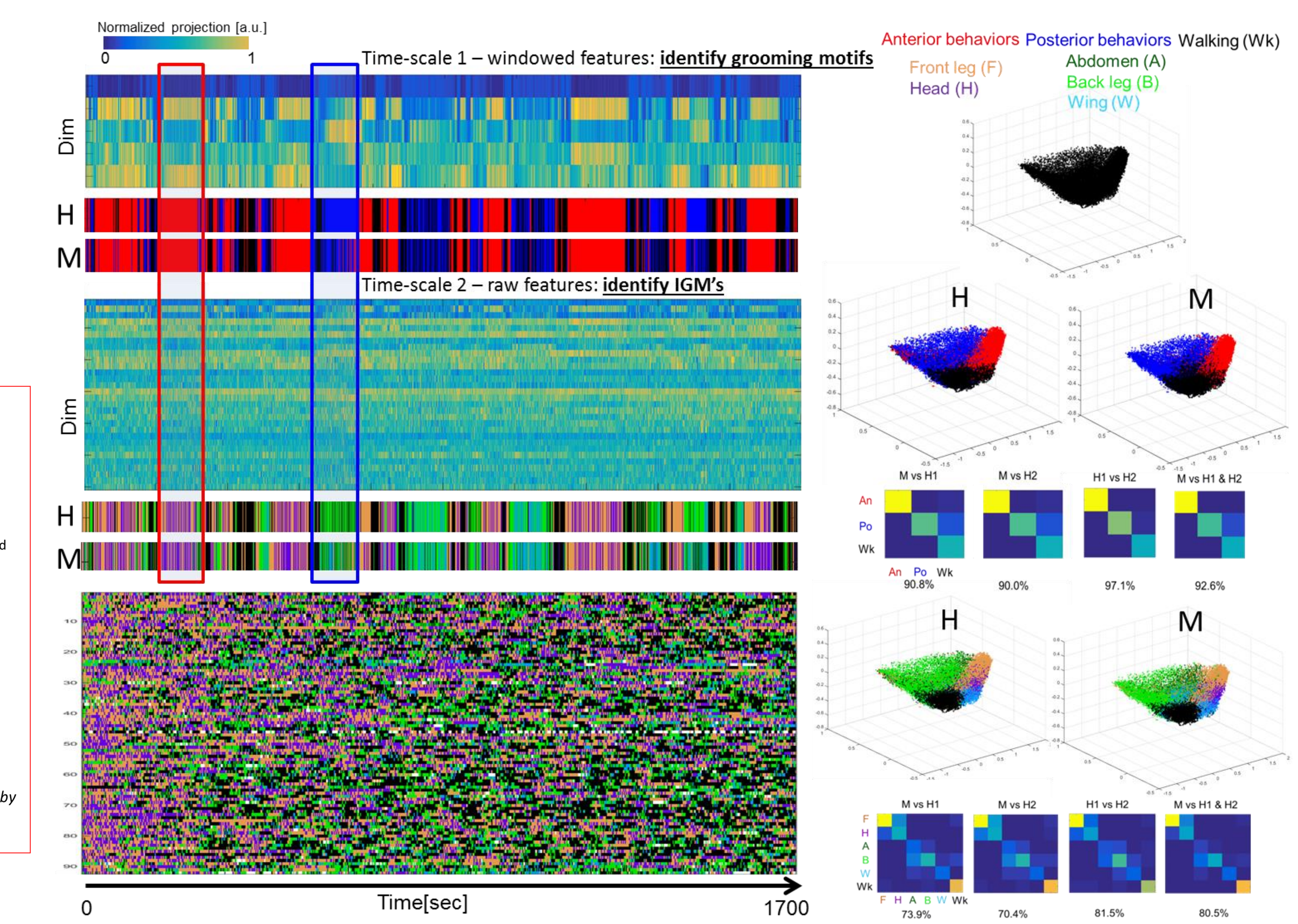
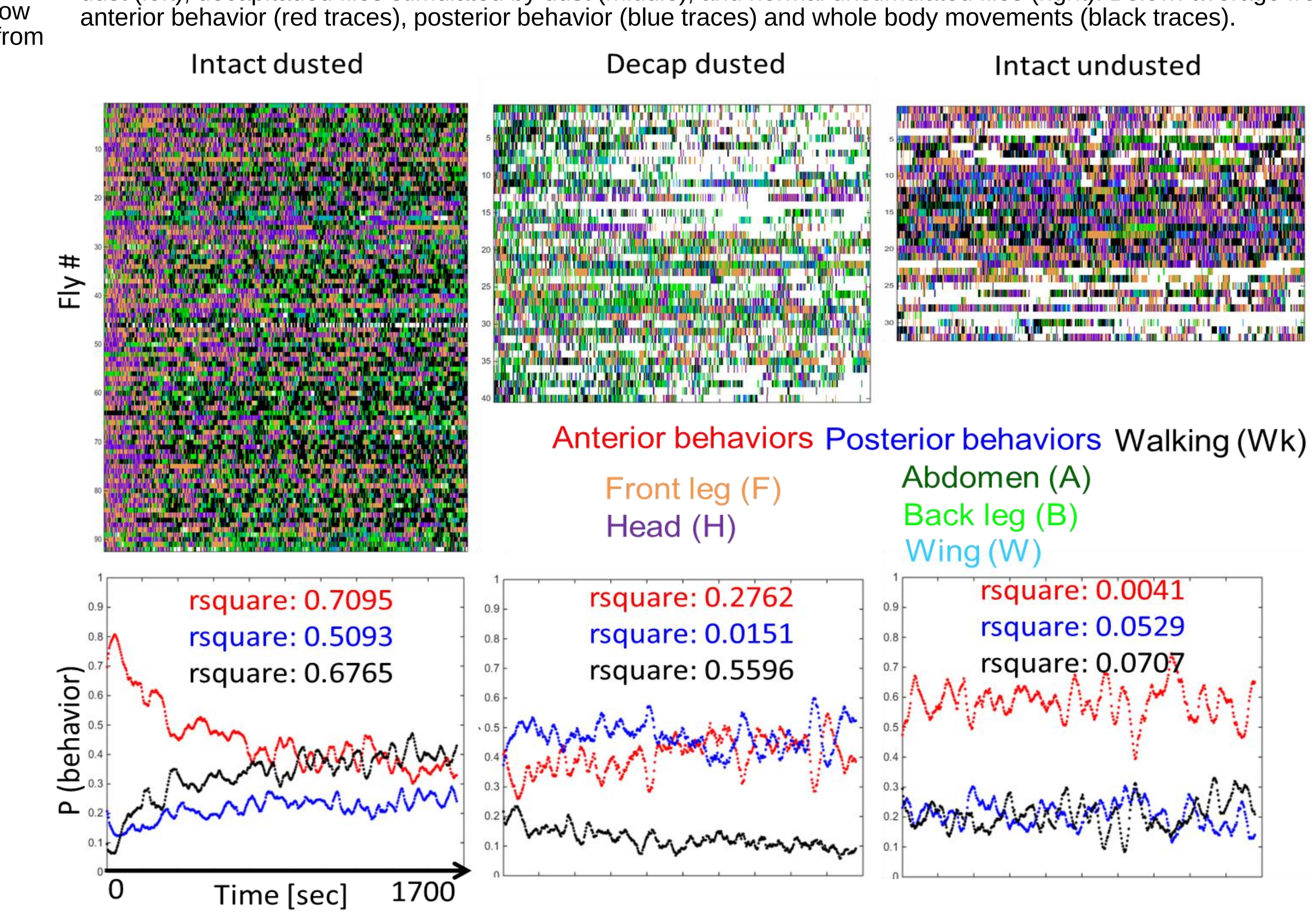


Figure 6 Results: comparing sequences

Population analysis of grooming behavior can be done with large sample sizes. Ethograms of 93 normal flies stimulated by dust (left), decapitated flies stimulated by dust (middle), and normal unstimulated flies (right). Below: average frequencies of anterior behavior (red traces), posterior behavior (blue traces) and whole body movements (black traces).



Conclusion: We can effectively use this robust behavior recognition method to analyze massive amounts of video data of freely walking flies across diverse experimental and recording conditions. The method may be expandable to other animal behaviors. The automatically recognized behavior is comparable to human-annotated behavior. The results can be used to study sequential decision-making in large numbers of animals, across varying recording conditions.

Acknowledgments: Funding: HHMI/Janelia; Data collection: Phuong Chung; spider data: Andrew Gordus